Title	Polymorphisms in the LAC12 gene explain lactose utilisation variability in Kluyveromyces marxianus strains
Authors	Varela, Javier A.;Montini, Noemi;Scully, Damhan;Van der Ploeg, Ralph;Oreb, Mislav;Boles, Eckhard;Hirota, Junya;Akada, Rinji;Hoshida, Hisashi;Morrissey, John P.
Publication date	2017-04-18
Original Citation	Varela, J. A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., Hirota, J., Akada, R., Hoshida, H. and Morrissey, J. P. (2017) 'Polymorphisms in the LAC12 gene explain lactose utilisation variability in Kluyveromyces marxianus strains', FEMS Yeast Research, 17(3), fox021 (13pp). doi:10.1093/femsyr/fox021
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1093/femsyr/fox021
Rights	© 2017, FEMS. All rights reserved. This is a pre-copyedited, author-produced version of an article accepted for publication in FEMS Yeast Research following peer review. The version of record Varela, J. A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., Hirota, J., Akada, R., Hoshida, H. and Morrissey, J. P. (2017) 'Polymorphisms in the LAC12 gene explain lactose utilisation variability in Kluyveromyces marxianus strains', FEMS Yeast Research, 17(3), fox021 (13pp). doi:10.1093/femsyr/fox021] is available online at: https://doi.org/10.1093/femsyr/fox021
Download date	2023-05-04 20:24:10
Item downloaded from	http://hdl.handle.net/10468/6273

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains

Javier A. Varela[1], Noemi Montini[1], Damhan Scully[1], Ralph Van der Ploeg[1], Mislav Oreb[2], Eckhard Boles[2], Junya Hirota[3], Rinji Akada[3,4,5], Hisashi Hoshida[3,4,5], and John P Morrissey[6] *

[1]School of Microbiology, University College Cork, Cork Ireland

[2]Institute of Molecular Biosciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[3]Department of Applied Chemistry, Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube 755-8611, Japan.

[4]Biomedical Engineering Center, Yamaguchi University, Ube 755-8611, Japan

[5]Research Center for Thermotolerant Microbial Resources, Yamaguchi University, Yamaguchi 753-8315, Japan

[6]School of Microbiology / Centre for Synthetic Biology and Biotechnology / Environmental Research Institute / APC Microbiome Institute, University College Cork, Cork T12YN60 Ireland

Manuscript for publication in: FEMS Yeast Research

Running title: *LAC12* SNPs in *K. marxianus*, sugar transporters

* To whom correspondence should be addressed. Tel.: +353 21 4902396; E-mail: J.Morrissey@ucc.ie

Keywords: Lactose utilisation, *LAC12*, *K. marxianus*.

Abstract

Kluyveromyces marxianus is a safe yeast used in the food and biotechnology sectors. One of the important traits that sets it apart from the familiar yeasts, *Saccharomyces cerevisiae*, is its capacity to grow using lactose as a carbon source. Like in its close relative, *Kluyveromyces lactis*, this requires lactose transport via a permease and intracellular hydrolysis of the disaccharide. Given the importance of the trait, it was intriguing that most, but not all, strains of *K. marxianus* are reported to consume lactose efficiently. In this study, primarily through heterologous expression in *S. cerevisiae* and *K. marxianus*, it was established that a single gene, *LAC12*, is responsible for lactose uptake in *K. marxianus*. Strains that failed to transport lactose showed variation in 13 amino acids in the Lac12p protein, rendering the protein non-functional for lactose transport. Genome analysis showed that the *LAC12* gene is present in 4 copies in the sub-telomeric regions of three different chromosomes but only the ancestral *LAC12* gene encodes a functional lactose transporter. Other copies of *LAC12* may be non-functional or have alternative substrates. The analysis raises some interesting questions regarding the evolution of sugar transporters in *K. marxianus*.

Introduction

Kluyveromyces marxianus is a yeast commonly isolated from food, fruit and plant material. It has frequently been isolated from fermented dairy products, such as cheese, yoghurt and kefir, where it is associated with production of pleasant flavours and other traits (Morrissey *et al*., 2015). It is closely related to *Kluyveromyces lactis*, itself considered to be a milk or dairy yeast (Lachance, 2007; Rodicio *et al*., 2013). Because of its history of association with food, *K. marxianus,* like *K. lactis*, has European Food Safety Authority (http://www.efsa.europa.eu/) Qualified Presumption of Safety (QPS) status and is a Generally Regarded as Safe (GRAS) organism. Despite the close relationship between *K. marxianus* and *K. lactis*, there are important phenotypic differences that have particular implications when considering the yeast for biotechnological applications (Fonseca *et al*., 2008; Lane and Morrissey, 2010). For instance, *K. marxianus* encodes an inulinase enzyme that gives the yeast the capacity to hydrolyse complex plant fructans (Arrizon *et al*., 2012, 2011), and it also has the capacity to grow at temperatures of 44 °C and above, both traits absent in *K. lactis* (Nonklang *et al*., 2009). *K. marxianus* is a respire-fermentative yeast and some strains are also reasonably good ethanol producers. Consequentially, there have been many studies seeking to exploit the beneficial traits of *K. marxianus* for bioethanol production. For example, ethanol has been produced from different substrates such as glucose (Castro and Roberto, 2014), lignocellulosic material (Goshima *et al*., 2013) and Jerusalem artichoke (Hu *et al*., 2012).

The capacity of *K. marxianus* and *K. lactis* to utilise lactose is one of the phenotypes that distinguishes them from *Saccharomyces cerevisiae,* which lacks this feature. Important biotechnology applications derive from this: namely, production of the enzyme β-galactosidase for the food industry, and remediation of whey by lactose removal (Rubio-Texeira, 2006). Whey is the waste material remaining from cheese manufacture and was long considered a problem until it was realised that whey permeate (deproteinised whey), which contains up to 5% residual lactose, can serve as a cheap substrate for industrial applications. As a result, *K. marxianus* is now used commercially for production of ethanol and yeast biomass from whey permeate (reviewed by Guimarães *et al*., 2010). These are, however, low value bulk products and the interest in yeast biotechnology is now to engineer strains to produce higher value products like flavours/aromas or low molecular weight bioactives (Morrissey *et al*., 2015). An ideal scenario from a biotechnology perspective is to be able to grow a yeast on a cheap substrate like whey permeate to produce high-value products and so the capacity to transport and metabolise lactose is very important.

Lactose utilisation in yeast involves transport of the sugar into the cell, intracellular hydrolysis of lactose into glucose and galactose, which are then metabolised through the glycolytic and Leloir pathways. The genetics of lactose assimilation have been intensively studied in *K. lactis*, where it is shown that the *LAC12 – LAC4* gene pair encodes a lactose permease and a β-galactosidase enzyme, respectively (Godecke *et al*., 1991). The Lac12p protein is a member of the major facilitator superfamily (MFS) and works via a

proton-symport mechanism. It is saturable at high substrate concentrations and is also reported to transport galactose. Therefore, it is considered a high-affinity lactose transporter and low-affinity galactose transporter in *K. lactis* (Riley *et al*., 1987). *K. lactis* also has at least one other transporter capable of transporting galactose but Lac12p is the only lactose transporter described (Baruffini *et al*., 2006). Notably, *Kluyveromyces* spp. lack the *GAL2* gene, which encodes the main galactose transporter in *S. cerevisiae* (Tschopp *et al*., 1986). *LAC12* and *LAC4* are divergently transcribed from a common promoter where key regulatory elements are located. This 2.8 kb *LAC12-LAC4* intergenic region contains several binding sites for Lac9p, which is the orthologue of the well-studied *S. cerevisiae* transcriptional activator, Gal4p (Godecke *et al*., 1991), as well as binding sites for the transcriptional repressor Mig1p. *LAC12* and *LAC4* are therefore strongly induced by lactose or galactose in the growth medium but potentially repressed by the presence of glucose.

The same genes, *LAC12* and *LAC4*, are present in *K. marxianus* and these are believed to account for the capacity of this yeast to utilise lactose. Several studies, however, suggest that efficient lactose-utilisation is not universally found in *K. marxianus*. For instance, differences in biomass production by *K. marxianus* strains grown in whey permeate have been reported (Grba *et al*., 2002). These findings are further supported by evidence suggesting that some *K. marxianus* strains lack lactose-proton symport activity and do not ferment lactose (Carvalho-Silva and Spencer-Martins, 1990). More recently, studies under defined media conditions reported that some strains either grew

very slowly, or not at all, when provided with lactose as the sole carbon source (Lane *et al*., 2011; Rocha *et al*., 2011). Publication of *K. marxianus* genome sequences has highlighted some interesting questions regarding lactose utilisation in *K. marxianus*. For example, it was reported that strain DMKU3-1042 encodes 4 genes that are highly homologous to the *LAC12* gene (Lertwattanasakul *et al*., 2015).

Given the importance of Lac12p for lactose transport and biotechnological applications of *K. marxianus*, we were interested in determining the reason for the phenotypic variability among a number of strains for this trait. The availability of draft and completed genome sequences enabled us to build a more complete view of the potential lactose transport genes in this yeast. Through a series of phenotypic, biochemical and genetic approaches, it was possible to determine that there is a single functional lactose transporter in *K. marxianus* and, in some strains, this protein no longer transports lactose into the cell. Apart from immediate relevance when engineering strains for applications, the work opens up some interesting questions regarding the evolution of yeast sugar transporters.

Materials and methods

Strains and culture conditions

The yeast strains used in study are listed in Table 1. Strains were typically cultured in YP medium (10 g/L yeast extract, 20 g/L bactopeptone) supplemented with 20 g/L glucose, lactose or raffinose. To determine growth kinetics and sugar transport, CBS 397 and CBS 6556 were grown on minimal media (MM) (Fonseca *et al*., 2007). For β - galactosidase measurements, cells were grown to mid exponential phase (A_{600} 2) on YPD and YPLac media. The *K. lactis* strain CBS 2359 was used as a control in this experiment. Enzymatic activity was measured as described by Ribeiro *et al*., 2007. *S. cerevisiae* EBY.VW4000 and BY4741 were used for heterologous expression and cellular localization experiments, respectively. EBY.VW4000 was grown in synthetic complete (SC) medium (1.7 g/L yeast nitrogen base, 5 g/L ammonium sulphate plus synthetic complete drop-out lacking L-leucine and uracil) (Formedium, Norfolk, United Kingdom). The growth medium was supplemented with 380 mg/L L-leucine and 76 mg/L uracil (final concentration) when required. Maltose or lactose was added as carbon source to a final concentration of 20 g/L. *S. cerevisiae* BY4741 was typically grown in YP medium supplemented with 20 g/L glucose or galactose and 200 µg/mL G418, when required for plasmid selection. Yeast strains were grown at 30 °C with 200 rpm agitation in a New Brunswick Innova 40/40R shaker (Eppendorf, Hamburg, Germany). Cloning experiments were performed using *Escherichia coli* DH5α. *E. coli* was routinely cultured in LB medium (5 g/L yeast extract, 10 g/L bactopeptone, 10 g/L NaCl) supplemented with 100 µg/mL ampicillin when required.

Analytic methods for sugar detection

Lactose concentration was determined by high-performance liquid chromatograph (HPLC). An Agilent 1200 HPLC system (Agilent Technologies, CA, USA) equipped with a refractive index detector and a REZEX 8 µL 8 % H^+ organic acid column (300 x 7.8 mm) (Phenomenex, CA, USA) was used to separate and detect the analytes. The column was eluted with 0.01 N H_2SO_4 at a flow rate of 0.6 mL/min. The temperature of the column was maintained at 65 °C. Lactose consumption was calculated by subtracting the lactose concentration at the end and the beginning of the experiment.

Sugar uptake assays to measure uptake of lactose and glucose were performed as previously described (Farwick *et al*., 2014). *K. marxianus* CBS 397 and CBS 6556 were grown on MM supplemented with glucose or lactose, as appropriate, and collected at an A_{600} of 2, which corresponds to mid exponential phase. Cells were centrifuged and then washed twice in cold uptake buffer (100 mM potassium phosphate, pH 6.5). Cell aliquots were equilibrated at 30 °C and then mixed with 100 mM radiolabelled glucose or lactose (American Radiolabeled Chemicals, MO, USA), for 15 seconds. Radioactivity was measured in a Wallac 1409 liquid scintillation counter. Counts from the unfiltered sample, referred as total counts, and from non-specific binding, were also determined. Intracellular sugar concentration was established by subtracting the non-specific binding counts to the filter counts and dividing this value by the total counts. Finally, uptake velocity was calculated by dividing the intracellular sugar concentration by the dry weight

and multiplying this value by the time of the assay. All experiments were carried out as 3 independent replicates.

Bioinformatic analysis

The unannotated genome sequence from the *K. marxianus* strain CBS 6556, also known as KCTC 17555, is publicly available (Jeong *et al*., 2012). Scaffold sequences (assembly GCA_000299195.2) were retrieved from the NCBI database and then submitted to the YGAP server (Yeast Genome Annotation Pipeline) for gene prediction (Proux-Wéra *et al*., 2012). Predicted proteins were then annotated using Blast2GO v 2.7.2 (Conesa *et al*., 2005). The genome of the *K. marxianus* strain DMKU3-1042, used for high-temperature ethanol fermentations, was used for genome comparisons (Lertwattanasakul *et al*., 2015; Nonklang *et al*., 2008). Clustering of the full set of genes of *K. marxianus* CBS 6556 and *K. marxianus* DMKU3-1042 (assembly GCA_001417885.1) was computed using OrthoMCL v 5 (Li *et al*., 2003).

To investigate sequence variation in the *LAC12* genes, the fragments of interest were amplified by PCR, sequenced and then multiple sequence alignments were created using AliView (Larsson, 2014). Maximum likelihood trees of the Lac12p sequences were computed using MEGA v 6.0.6 (Tamura *et al*., 2013). The CBS 6556 and CBS 397 *LAC12* sequences can be found in the EBI site under the ID numbers LT708111-LT708118

Functional expression of *LAC12* genes

All the plasmids used in this study are listed in Table 2. All yeast transformations were performed using the standard Lithium Acetate

procedure (Gietz and Schiestl, 2007). Plasmids for functional expression of *LAC12* genes in *S. cerevisiae* were constructed using the pGREG-505 backbone (Jansen *et al*., 2005). The *GAL1* promoter in pGREG-505 was replaced with the constitutive *TEF1* promoter. The *TEF1* sequence was amplified from BY4741, digested with NotI/SacI and cloned into the similarly cut pGREG-505 using T4 DNA ligase (New England Biolabs Inc., MA, USA) to yield plasmid pGREG-505-TEF1. The *LAC12* genes were cloned into pGREG505-TEF1 by *in vivo* recombination in *S. cerevisiae*. The four *LAC12* open reading frames were amplified by PCR from CBS 397 and CBS 6556 using Phusion Polymerase (New England Biolabs Inc., MA, USA). Then, 1 µL of each product was mixed with 200 ng of *SalI*-digested pGREG505-TEF1 and transformed into *S. cerevisiae* EBY.VW4000**.** Cells were plated on SC media lacking leucine. Colonies were screened by replica plating in –His media and colony PCR. Plasmids were recovered from yeast cultures and transformed into *E. coli* as described elsewhere (Singh and Weil, 2002).

To express the β-galactosidase gene required for lactose hydrolysis, the pGREG506-TEF1-*LAC4* was constructed by recombinational cloning. Briefly, the *TEF1* promoter from *S. cerevisiae*, the *LAC4* gene from *K. marxianus* CBS 397 and the pGREG506 plasmid, containing the uracil maker, were PCR amplified using primers designed for *in vivo* assembly (Table 3). As mentioned previously, 1 µL of each product was transformed into *S. cerevisiae* EBY.VW4000 and the resulting plasmid was recovered. Finally, this plasmid was transformed into the EBY.VW4000 strain containing the

different *LAC12* plasmids. Cells were plated in SC media lacking uracil and leucine.

The *LAC12* PCR products were also cloned into the pGREG574 plasmid to obtain in frame N-terminal GPF fusions. In this case, the *S. cerevisiae* BY4741 strain was used. *S. cerevisiae* BY4741 strains transformed with the pGREG574 – *LAC12* plasmids were grown in inducing YPGal medium to an A_{600} of 2. Cells were collected and washed twice with sterile water. Suspensions of 10 µL were then spread into glass slides. Samples were visualized using a Zeiss LSM 5 laser scanning confocal microscope. For detection of GFP fluorescence, samples were excited with laser light at 488 nm and emission light acquired using a bandpass emission filter of 505-550 nm. Samples were acquired and analysed using Zen 2008 SP2 software.

For expression in *K. marxianus*, the *KmLAC12* genes were cloned into YCp11256 (Fig. S1) under the *KmCDC19* promoter by using functional marker selection (Hoshida *et al*., 2014). The vector DNA was linearized and amplified with primers *Km*CDC19-1c and *URA3*+771c. *KmLAC12* genes from NCYC 1429 and DMKU3-1042 were amplified with primers shown in Table 3. The *URA3* C-terminal sequence was added to the resulting *KmLAC12* gene fragments by PCR using each +1 primer and *KmURA3*+772786TGA+3CG9. The linearized vector DNA and one of *KmLAC12* gene fragments were simply mixed and used for transformation of RAK3605, a *ura3-1* derivative of DMKU3-1042 (Nonklang *et al*., 2008). *In vivo* assembly using the NHEJ repair mechanism joined the fragments end to end, recreating an intact *URA3* gene

and allowing selection of transformants on SC media lacking uracil. Correct cloning of each *KmLAC12* gene was confirmed by PCR and sequencing.

Gene expression analysis

K. marxianus strains were grown in YPD, YPLac and YRaff to an A_{600} of 2. Cells were then recovered by centrifugation and RNA was extracted using the Ambion RiboPure RNA Purification Kit for Yeasts (Ambion, Austin, TX, USA). The extracted RNA was quantified using a Nanodrop spectrophotometer. cDNA was synthesized from 800 ng of RNA, using the ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs Inc., MA, USA). cDNA samples were diluted by adding 30 µL of nuclease-free water, resulting in a final volume of 50 µL per cDNA sample. Aliquots were then diluted 1/10 and stored at -20 °C. Real-time quantitative PCR (qPCR) was used to measure gene expression. cDNA samples were amplified by using LightCycler 480 SYBR Green I Master qPCR and ran in a LightCycler 480 real-time PCR machine. Primers were design to amplify the *ACT1, ALG9*, *LAC4,* and *LAC12* genes (Table 3). Standard curves were performed to check for amplification efficiency with all primers showing > 90 % efficiency. Additionally, melting curves were carried out to check reaction specificity and to ensure that each primer pair produced a single amplicon. *ALG9* (coding for alpha-1,2-mannosyltransferase) and *ACT1* (coding for actin) were used as reference genes, as both these genes are commonly used reference genes involved in cell processes, unrelated to lactose metabolism (Teste *et al*. 2009). To compare expression of the *LAC12* genes in control media varied between strains, the data was normalised against *ACT1* expression. Data was further

calculated as $2^{-\Delta Ct}$ as described by (Pfaffl, 2001). *ALG9* was used as an internal control to ensure that the *ACT1* expression was constant in all samples.

Genetic crosses and spore analysis

RAK 3605 and RAK 4072 (DMKU3-1042 *ura3-1*, and NCYC1429 *ura5-2*, respectively, were crossed on 2% glucose medium and incubated at 30ºC for one day (Yarimizu *et al*., 2013). Yeast cells were replica-plated on a MM plate and incubated for two days to obtain zygotes. For sporulation, zygotes were suspended in 4 mL sterilized water at a cell concentration A_{600} 0.1. After two-day incubation, cells were spread on MM supplemented with uracil and 1 mg/mL 5-Fluoroorotic acid. One hundred ninety-two colonies were randomly selected and used for growth tests in YPLac medium.

Results

K. marxianus strains can be classified according to their lactose consumption capacities

Although the capacity to utilise lactose is considered to be a typical trait of *K. marxianus*, several studies have found this trait to be variable between strains (Carvalho-Silva and Spencer-Martins, 1990; Grba *et al*., 2002; Lane *et al*., 2011; Rocha *et al*., 2011). In order to establish the degree of variability, lactose utilisation was evaluated in a set of 15 *K. marxianus* strains derived from culture collections and originating from different niches (Table 1). Strains were grown on YPLac and growth was assessed after 11h (Fig 1, open bars). In the same experiment, the amount of lactose consumed after 11h of growth was determined (Fig 1, solid bars). In both cases, there was a very clear bimodal distribution with 6 strains (CBS 397, CBS 608, NCYC 179, CBS 5795, CBS 6432 and CBS 1555) growing to a higher density (~A_{600} 1.8) and consuming a higher amount of lactose and 9 strains (CBS 6556, CBS 5670, CBS 4857, CBS 712, CBS 7894, CBS 7858, CBS 1596, CBS 2233 and CBS 4354) growing poorly (A_{600} < 1) and consuming low amounts of lactose. The precise overlap between growth and lactose consumption in YPLac suggests that strains can be classified as either poor or good consumers of lactose based on growth patterns.

CBS 397 transport lactose more efficiently than CBS 6556

To study this trait in more detail, two strains, CBS397 and CBS 6556, which are among the most studied in the field, were selected as representatives of the good and poor lactose consumers (Etschmann *et al*., 2003; Fonseca *et al*., 2013, 2007; Jeong *et al*., 2012; Wang and Bajpai, 1997). First, to test whether the differences observed were exclusive to lactose consumption, the yeasts were grown in MM glucose (Fig. 2A). Both strains grew equally well when grown in this media reaching stationary phase after 11-12 hours. Sugar consumption was also monitored showing that CBS 397 and CBS 6556 utilised all the glucose in the media. In contrast, when the strains were cultured in MM lactose (Fig. 2B), CBS 397 grew at a similar rate as when using glucose as a carbon source but CBS 6556 exhibited a dramatic reduction in growth rate. Lactose consumption analysis showed that CBS 397 utilised lactose efficiently whereas CBS 6556 barely consumed any sugar. Since there was some growth of CBS 6556 on MM with 2% lactose with little apparent use of lactose, growth on the lower concentration of 0.2% lactose was assessed (Fig. 2C). Although overall yields (as measured by A_{600}) were lower, a similar pattern of growth on 2% lactose was observed. In this case, it was possible to observe low levels of lactose utilisation by CBS 6556 indicating that the strain can utilise lactose, but at a much lower rate than CBS 397. In control experiments with MM lacking any sugar, neither strain could grow, thereby confirming that all growth observed relied on lactose as the carbon source (data not shown). As a key step in lactose utilisation is the efficient hydrolysis of this sugar into glucose and galactose, β - galactosidase activity was measured in the strains to test for possible differences. Overall,

low activity was detected when the strains were grown in glucose and high levels when grown in lactose (Fig 2D). Interestingly, CBS 6556 showed higher levels of activity then CBS 397 but lower than the *K. lactis* strain CBS 2359 suggesting that differences in β - galactosidase are not accountable for the growth defect exhibited by CBS 6556. Next, to establish whether the variation in lactose consumption was due to differences in lactose transport efficiency, sugar uptake assays were conducted using radiolabelled substrates (Fig. 2E). Glucose and lactose were assayed separately using saturating concentrations and maximum velocities (Vmax) were calculated accordingly. The strains did not show significant differences in glucose transport (Fig. 2E, black bars) but the Vmax for lactose transport was 10 times higher in CBS 397 than in CBS 6556 (Fig. 2E, white bars). These results indicate that the phenotypic differences observed between the strains can be explained by changes in lactose transport efficiency.

K. marxianus strains carry four copies of *LAC12*

In the sister species, *K. lactis*, lactose is transported by the Lac12p permease, therefore one possible explanation for the differences could have been deletion or inactivation of the CBS 6556 *LAC12* gene. To investigate this possibility, the published genome sequence of CBS 6556 was interrogated using tBLASTn and the *K. lactis* Lac12p. This identified 3 homologous sequences, all of which were automatically annotated as *LAC12* by Blast2GO. As Lertwattanasakul *et al*., 2015 showed the presence of 4 putative *LAC12* genes in the DMKU3-1042 genome, the CBS 6556 and DMKU3-1042 genomes were compared. Three of the *LAC12* genes were present in both

strains but one *LAC12* gene, located in the subtelomeric region of chromosome 8, appeared to be absent in CBS 6556. Comparative genome analysis, however, revealed that a 50 kb section is missing from the CBS 6556 assembly and PCR confirmed the presence of the 4th copy of *LAC12* in CBS 6556 (data not shown). Thus, both CBS 6556 and DMKU3-1042 carry 4 very similar copies of the *LAC12* gene in virtually identical genomic contexts (Fig. 3) One of the *LAC12* genes, located in the subtelomeric region of chromosome 3, shares the *K. lactis LAC12-LAC4* locus topology and is also the closest in sequence to *K. lactis LAC12* and is therefore designated *LAC12* (*KmLAC12).* The homologous sequences, designated *LAC12-2, LAC12-3 and LAC12-4,* are found on the opposite extreme of chromosome 3 (to *KmLAC12*), on chromosome 2, and on chromosome 8, respectively. Interestingly, in CBS 6556, *LAC12-2* also displays the *K. lactis LAC12-LAC4* locus organisation except that the *LAC4* sequence at this locus has in-frame stop codons, indicating that it is a non-functional pseudogene. In DMKU3-1042, the *LAC4* pseudogene is absent and *LAC12-2* is flanked at the 5' end by a putative oxidoreductase. With that exception, the genomic context of the *LAC12* genes is similar in both CBS 6556 and DMKU3-1042. More extensive analysis in other *K. marxianus* strains revealed that all tested strains carry the 4 *LAC12* genes but in some (CBS 6556 and DMKU3-1042), the *LAC12-4* gene is a pseudogene whereas in others, the sequence appears to be capable of encoding a functional protein (Fig. S2). Analysis by PCR and sequencing established that CBS 397 carries the four *LAC12* genes, including a putative functional *LAC12-4* gene. The presence of the *LAC4* pseudo gene, next to *LAC12-2*, was also confirmed in this strain.

A single gene is responsible for the differences in lactose consumption

The presence of multiple *LAC12* copies suggested that the lactose consumption differences between *K. marxianus* strains could be caused by the activity of multiple genes. To test whether this was the case or whether a single gene was responsible for the phenotypic variation, an analysis of a genetic cross between representatives of good and poor lactose consumers was carried out. Because specific markers were required for this analysis, auxotrophic derivatives of DMKU3–1042, which displays the CBS 6556 poor lactose consumption phenotype, and NCYC 1429, which represents the CBS 397 phenotype (Fig. S3), were used. For this experiment, the two strains were mated, the diploid sporulated, and random spore analysis was performed, scoring for the lactose utilisation phenotype (Fig. 4). The diploid had the lactose-utilisation (NCYC 1429) phenotype and all progeny showed either one of the parental phenotypes and the absence of intermediate phenotypes. These genetic data indicate that a single dominant gene is responsible for the differentiation of phenotypes.

KmLAC12 is induced by lactose in both good and poor lactose-utilising strains

Since it is known in *K. lactis* that expression of the *LAC12* and *LAC4* genes is strongly induced by lactose, RTqPCR analyses were performed in strains CBS 397 and CBS 6556 (Fig. 5). Strains were grown to mid-exponential phase on YP with 2% glucose, raffinose or lactose as the carbon source and expression compared to the *ACT1* gene assessed by RTqPCR. There were marked differences between the *LAC12* genes, with *KmLAC12* expressed to a much higher level than the other *LAC12* genes. There was also strong

induction on lactose compared to raffinose or glucose, possibly as a consequence of the presence of putative Gal4p binding sites in the *LAC12-LAC4* intergenic region (Fig. 3). *LAC12-2, LAC12-3* and *LAC12-4* all showed lactose induction to some degree but the level of expression was low compared to *KmLAC12*. For example, in the case of *LAC12-2* and *LAC12-4*, induced expression was only at the level of basal expression of *KmLAC12*. It was significant that there was no substantial difference in expression patterns between CBS 6556 (open bars) and CBS 397 (solid bars), suggesting that expression differences do not explain the lactose utilisation phenotype. Expression of *LAC4* is also strongly induced but comparable in both strains, consistent with earlier data showing that differences in β-galactosidase activity are not responsible for the different phenotypes of the good and poor lactose utilising strains (Fig. 2D).

Functional differences in the *Km*Lac12p protein are responsible for the variable lactose-utilisation phenotype.

Since neither comparative genomics nor expression analysis identified any differences that would account for why lactose transport is efficient in CBS 397, NCYC 1429 and other strains, but poor in CBS 6556, DMKU-1042 and other strains, heterologous expression of the *LAC12* genes was carried out. First, each of the 4 genes from each strain was expressed in *S. cerevisiae* to assess lactose transport (Fig. 6A). The host strain, *S. cerevisiae* EBY.VW4000 lacks hexose transporters and was used to avoid growth on hexoses which might arise from impurities in the lactose or as a consequence of unspecific lactose hydrolysis. However, this strain can grow on maltose

using the Mal2p maltose transporter (Wieczorke *et al*., 1999). *S. cerevisiae* strains expressing the different *LAC12* genes (and the β-galactosidase gene, *LAC4)* were washed, diluted and plated onto different plates with either maltose or lactose as the carbon source. All strains grew on the maltose control plate, but when growth in lactose media was examined, it was found that only the CBS 397 version of *KmLAC12* conferred growth. This indicates that, of the 4 *LAC12* genes in CBS 397, only *KmLAC12* encodes a functional lactose transporter. Furthermore, it demonstrates that the same protein from CBS 6556 is not able to transport lactose. To rule out a problem of expression or protein localisation of the CBS 6556 version of *Km*Lac12p, an in-frame fusion of CBS 6556 *Km*Lac12p to GFP was made. Fluorescence microscopy showed that the fusion protein correctly localised to the cell membrane (Fig. S4), suggesting that CBS 397 and CBS 6556 genuinely have different variants of *KmLAC12*.

This data suggested that expression of the *KmLAC12* allele from a good lactose-utilising strain in a poor lactose-utilising strain would be sufficient to change the strain phenotype. To test this hypothesis, the two versions of the gene were expressed in DMKU3-1042, a poor utiliser (Fig. 6B). In this case, expression of the NCYC 1429 *KmLAC12* gene (identical in sequence to CBS 397) conferred growth on lactose, whereas expression of the DMKU 3-1042 gene (identical in sequence to CBS 6556) had no effect. Furthermore, none of other three *LAC12* genes (*LAC12-2, LAC12-3* or *LAC12-4*) cloned from NCYC 1429 conferred growth, consistent with the conclusion that in

good lactose utilisers like NCYC 1429 and CBS 397, only *KmLAC12* encodes a lactose transporter.

Changes in amino acid sequence are responsible for lactose consumption variability

The amino acid sequence of the CBS 397 / NCYC 1429, the CBS 6556 / DMKU3-1042, and the K. lactis CBS 2359 KmLac12p proteins were compared to identify the changes that could give rise to the functional variation. This revealed 19 variant amino acids across the transporter (Fig. 7). To add to the resolution, the same gene was fully sequenced in the 13 strains shown in Figure 1. With the exception of CBS 1596, in which there was a deletion rendering *KmLAC12* a pseudogene, full protein sequences were obtained for all strains. Multiple sequence alignment (Fig S5) showed that there were two sequence variants of *Km*Lac12p. These are differentiated by changes in 13 amino acids (Fig 7, red shading). The distribution of these residues did not seem to show any particular pattern regarding the potential protein structure: changes were found in transmembrane domains and cytoplasmic regions alike. The *K. lactis* strain CBS 2359, known to be a good lactose consumer, also shared the CBS 397 amino acid residues that define the good lactose consumption allele.

To gain an insight into the evolution of the *LAC12* genes, protein sequences from CBS 6556 and CBS 397 were used to generate a phylogenetic tree (Fig. 8). As shown, the *Km*Lac12p sequences form a distinctive cluster that includes the *K. lactis* Lac12p transporter. Within this group the CBS 397

protein is more similar to the *K. lactis* sequence than to the CBS 6556 one. The remainder of the Lac12p sequences are located in another branch of the tree. Here, the Lac12-2p and Lac12-4p transporters are more closely related to each other than to Lac12-3p. The putative low-affinity galactose transporter *Km*Hgt1p (Baruffini *et al*., 2006) forms a distinct branch and is more related to the *S. cerevisiae* hexose transporters Hxt1p (glucose) and Gal2p (galactose).

Discussion

Transport and metabolism of lactose is required for many biotechnological applications of *K. marxianus* so it is important to understand the basis of this trait. The basic mechanism of transport via the Lac12p permease and hydrolysis by the intracellular Lac4p β-galactosidase in *Kluyveromyces* species has been known for some time from studies with *K. lactis* (Godecke *et al*., 1991). The presence of these genes is conserved in *K. marxianus* so the variable growth of different strains on lactose-containing medium was a conundrum. Further uncertainty came from the observation from genome sequencing projects that *K. marxianus* carries several genes that are homologous to *LAC12* and therefore, are also potential lactose transporters (Lertwattanasakul *et al*., 2015). In this study, it is unequivocally demonstrated that the main transporter responsible for uptake of lactose by *K. marxianus* is the *KmLAC12* gene that is located on the left arm of chromosome 3, transcribed divergently from the *LAC4* gene. As is the case with *K. lactis,* the intergenic region between these genes has several putative binding sites for the Gal4p transcription factor and both *LAC12* and *LAC4* are very strongly induced by lactose and show some evidence of glucose repression. The *LAC12* and *LAC4* nucleotide sequences show a high degree of conservation with those genes from *K. lactis* and it is clear that this *LAC12-LAC4* locus shares a common evolutionary origin in *K. lactis* and *K. marxianus* and is most likely ancestral to the speciation.

Since comparison with *K. lactis* indicates that the *K. marxianus* locus on the left art of chromosome 3 is ancestral, it follows that the other copies of *LAC12*

that are present in *K. marxianus* arose by gene duplication events. The location of all 4 homologous *LAC12* genes in sub-telomeric chromosome regions immediately suggests the mechanism of duplication. Yeast subtelomeric regions have been shown to be subject to a high rate of duplication and recombination (Barton *et al*., 2008). Indeed, in *K. lactis* 7 out of 12 subtelomeres contain a duplicated 9 kb region that contains several duplicated genes (Fairhead and Dujon, 2006). Our analysis of other chromosomes in *K. marxianus* found that several chromosomes have duplicated regions (Varela and Morrissey, unpublished data). In the case of chromosome 3, the *LAC12-LAC4* locus became duplicated on the right sub-telomeric region of chromosome 3 (*LAC12-2 – LAC4).* From comparison of sequenced *K. marxianus* strains and analysis of some strains by targeted PCR, it is seen that no strains retained a functional *LAC4* gene at this locus – in all cases it has either acquired deletion/frameshift mutations or it is completely absent, presumably lost by a deletion event. In contrast, the *LAC12-2* gene retains the coding capacity for an intact protein, albeit one that has lost the capacity to transport lactose. Phylogenetic analysis indicates that *LAC12-3* and *LAC12-4* probably arose from duplications of *LAC12-2* (Figure 9) so it is not surprising that these proteins also do not encode functional lactose transporters. In some strains (e.g. CBS 6556 and DMKU3-1042), *LAC12*-4 is a pseudogene, whereas in others (e.g. CBS 397, NCYC 1429) it still has the potential to encode a functional protein. The variation seen between strains suggests that diversification in the sub-telomeric regions is an ongoing process. More broadly, sub-telomeric regions have been shown to be hot spots of gene evolution and functional divergence (Anderson *et al*., 2015).

Rapid evolution of subtelomeric gene families has been shown in the *MAL* genes, required for maltose consumption in *S. cerevisiae* (Brown *et al*., 2010). Moreover, it was recently demonstrated that certain duplicated hexose transporter family members in *S. cerevisiae* (Hxt13 and Hxt15-Hxt17; all encoded in subtelomeric regions) have lost their ability to transport hexoses and evolved as hexitol transporters (Jordan *et al*., 2016).

As well as establishing that *KmLAC12* encodes the main *K. marxianus* lactose transporter, the work reported in this study determined that the reason why CBS 6556, DMKU3-1042 and other strains fail to grow efficiently on lactose is due to variation in that gene. More specifically, the *KmLAC12* allele in those strains has sufficiently diverged from the ancestral sequence to no longer encode a lactose transporter. In one poor lactose-utilising strain tested (CBS 1596), there was a deletion rendering the gene non-functional, but in all other cases, it is still possibly encoding an intact protein. Analysis of the amino acid sequence of 12 strains determined that there are 13 amino acid variations that distinguish a functional / non-functional lactose transporter. It remains to be established whether any single amino acid change is sufficient for this effect, or whether a combination of some or all of the 13 changes is required to abolish lactose transport. It has been previously shown that lactose permease specificity can be modified in *E. coli* (Varela *et al*., 1997). For example, it has been shown that chemical modification of a single cysteine residue in the lactose permease LacY could change the transporter into a galactose specific transporter (Guan *et al*., 2002). Although some of the studied amino acid changes in *E. coli* could potentially be applied to the *K. marxianus*

transporters, the *Km*Lac12p and LacY are only 33 % identical making the comparison difficult.

Comparative analysis of the amino acid sequence of each of the Lac12p proteins between CBS 397 and CBS 6556 is intriguing. There is very little variation between Lac12-2p (99 % identity) or Lac12-3p (100 % identity), but more variation between Lac12-4p (97 % identity) and *Km*Lac12p (97 % identity). In contrast, the Lac12p from CBS 397 and *K. lactis* are 99 % identical. One could speculate that the CBS 6556 *KmLAC12* gene has undergone selection to transport a sugar other than lactose. In this regard, it is notable that unlike CBS 397 (yoghurt), both CBS 6556 (pozol) and DMKU 3-1042 (bioethanol) were isolated from non-dairy environments and would not necessarily have a requirement to transport lactose. Since the substrate from which many of the strains used in this study were isolated is not known, it is not possible to extrapolate this observation. Nor is it known what this alternative substrate would be, but our analysis indicates that it is not galactose (Varela and Morrissey, unpublished). If this suggestion is correct, however, it seems peculiar that the β–galactosidase activity remains in yeast strains that have lost the capacity to utilise lactose. Thus, there are certainly still unanswered questions. It must also be considered that proteins in this family can also transport other organic and inorganic small molecules, so there is also the possibility of a non-sugar substrate. In any case, the phenotypic variability observed across strains indicates that lactose utilisation is not universal in *K. marxianus.* Although the original classification of the species assumed this trait to be present in all strains, our data supports the

idea that sequence-based methods are more robust than phenotypic assays for classifying yeasts.

Much of what is known about sugar transporters in eukaryotes has come from research with *S. cerevisiae*. In particular, that yeast has an impressive repertoire of high and low affinity glucose sensors and transporters that play a key role in the ecology and physiology of the organism (Ozcan and Johnston, 1999). Furthermore, research on maltose transport in brewing strains have been a model for how selection can give rise to gene families (Jespersen *et al*., 1999). There has been a perception that domestication had created some unique evolutionary pressures for *S. cerevisiae* and other yeasts will not show similar complexity. The findings of this study and others (Lertwattanasakul *et al*., 2015) indicate that this, in fact, is not the case. *K. marxianus* has a large family of related sugar transporters and decipering their regulatory and metabolic networks is a key challenge to aid in the exploitation of *K. marxianus* as a cell factory yeast for biotechnology.

Acknowledgements

J. V. is supported by the YEASTCELL Marie-Curie ITN project (REA Grant No. 606795) under the EU's Seventh Framework Programme for Research (FP7). Technical assistance of Dan Walsh with HPLC analysis, and Paddy O' Reilly with Microbiology work, is greatly acknowledged. This study was supported in part by the Advanced Low Carbon Technology R&D Program (JST, Japan).

Conflict of interest

The authors declare no conflict of interest

References

Anderson, M.Z., Wigen, L.J., Burrack, L.S., Berman, J., 2015. Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*. Genetics 200, 907–919. doi:10.1534/genetics.115.177451

Arrizon, J., Morel, S., Gschaedler, A., Monsan, P., 2012. Fructanase and fructosyltransferase activity of non-Saccharomyces yeasts isolated from fermenting musts of Mezcal. Bioresour. Technol. 110, 560–565. doi:10.1016/j.biortech.2012.01.112

Arrizon, J., Morel, S., Gschaedler, A., Monsan, P., 2011. Purification and substrate specificities of a fructanase from *Kluyveromyces marxianus* isolated from the fermentation process of Mezcal. Bioresour. Technol. 102, 3298–3303. doi:10.1016/j.biortech.2010.10.071

Barton, A.B., Pekosz, M.R., Kurvathi, R.S., Kaback, D.B., 2008. Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. Genetics 179, 1221–35. doi:10.1534/genetics.107.083493

Baruffini, E., Goffrini, P., Donnini, C., Lodi, T., 2006. Galactose transport in *Kluyveromyces lactis*: major role of the glucose permease Hgt1. FEMS Yeast Res. 6, 1235–42. doi:10.1111/j.1567-1364.2006.00107.x

Brown, C.A., Murray, A.W., Verstrepen, K.J., 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr. Biol. 20, 895–903. doi:10.1016/j.cub.2010.04.027

Carvalho-Silva, M., Spencer-Martins, I., 1990. Modes of lactose uptake in the yeast species *Kluyveromyces marxianus*. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 57, 77–81. doi:10.1007/BF00403158

Castro, R.C.A., Roberto, I.C., 2014. Selection of a Thermotolerant *Kluyveromyces marxianus* Strain with Potential Application for Cellulosic Ethanol Production by Simultaneous Saccharification and Fermentation. Appl. Biochem. Biotechnol. 172, 1553–1564. doi:10.1007/s12010-013-0612-5

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. doi:10.1093/bioinformatics/bti610

Etschmann, M.M., Sell, D., Schrader, J., 2003. Screening of yeasts for the production of the aroma compound 2-phenylethanol in a molasses-based medium. Biotechnol. Lett. 25, 531–6.

Fairhead, C., Dujon, B., 2006. Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. FEMS Yeast Res. 6, 428–441. doi:10.1111/j.1567-1364.2006.00033.x

Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., Boles, E., 2014. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. Proc. Natl. Acad. Sci. U. S. A. 111, 5159–64. doi:10.1073/pnas.1323464111

Fonseca, G.G., de Carvalho, N.M.B., Gombert, A.K., 2013. Growth of the yeast *Kluyveromyces marxianus* CBS 6556 on different sugar combinations as sole carbon and energy source. Appl. Microbiol. Biotechnol. 97, 5055–5067. doi:10.1007/s00253-013-4748-6

Fonseca, G.G., Gombert, A.K., Heinzle, E., Wittmann, C., 2007. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with

glucose as the sole carbon source. FEMS Yeast Res. 7, 422–35. doi:10.1111/j.1567-1364.2006.00192.x

Fonseca, G.G., Heinzle, E., Wittmann, C., Gombert, A.K., 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl. Microbiol. Biotechnol. 79, 339–354. doi:10.1007/s00253-008-1458-6

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34. doi:10.1038/nprot.2007.13

Godecke, A., Zachariae, W., Arvanitidis, A., Breunig, K.D., 1991. Coregulation of the Kluyveromyces-Lactis Lactose Permease and Beta-Galactosidase Genes Is Achieved by Interaction of Multiple Lac9 Binding-Sites in A 2.6-Kbp Divergent Promoter. Nucleic Acids Res. 19, 5351–5358.

Goshima, T., Tsuji, M., Inoue, H., Yano, S., Hoshino, T., Matsushika, A., 2013. Bioethanol Production from Lignocellulosic Biomass by a Novel *Kluyveromyces marxianus* Strain. Biosci. Biotechnol. Biochem. 77, 1505–1510. doi:10.1271/bbb.130173

Grba, S., Stehlik-Tomas, V., Stanzer, D., Vahcic, N., Skrlin, A., 2002. Selection of yeast strain *Kluyveromyces marxianus* for alcohol and biomass production on whey. Chem. Biochem. Eng. Q. 16, 13–16.

Guan, L., Sahin-Tó th, M., Ronald Kaback, H., 2002. Changing the lactose permease of *Escherichia coli* into a galactose-specific symporter.

Guimrães, P.M.R., Teixeira, J.A., Domingues, L., 2010. Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. Biotechnol. Adv. 28, 375–384. doi:10.1016/j.biotechadv.2010.02.002

Hoshida, H., Murakami, N., Suzuki, A., Tamura, R., Asakawa, J., Abdel-Banat, B.M.A., Nonklang, S., Nakamura, M., Akada, R., 2014. Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*. Yeast 31, 29–46. doi:10.1002/yea.2993

Hu, N., Yuan, B., Sun, J., Wang, S.-A., Li, F.-L., 2012. Thermotolerant Kluyveromyces marxianus and *Saccharomyces cerevisiae* strains representing potentials for bioethanol production from Jerusalem artichoke by consolidated bioprocessing. Appl. Microbiol. Biotechnol. 95, 1359–1368. doi:10.1007/s00253-012-4240-8

Jansen, G., Wu, C., Schade, B., Thomas, D.Y., Whiteway, M., 2005. Drag&Drop cloning in yeast. Gene 344, 43–51. doi:10.1016/j.gene.2004.10.016

Jeong, H., Lee, D.-H., Kim, S.H., Kim, H.-J., Lee, K., Song, J.Y., Kim, B.K., Sung, B.H., Park, J.C., Sohn, J.H., Koo, H.M., Kim, J.F., 2012. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. marxianus KCTC 17555. Eukaryot. Cell 11, 1584–5. doi:10.1128/EC.00260-12

Jespersen, L., Cesar, L.B., Meaden, P.G., Jakobsen, M., 1999. Multiple alpha-glucoside transporter genes in brewer's yeast. Appl. Environ. Microbiol. 65, 450–6.

Lachance, M.-A., 2007. Current status of *Kluyveromyces* systematics. FEMS Yeast Res. 7, 642–645. doi:10.1111/j.1567-1364.2006.00197.x

Lane, M.M., Burke, N., Karreman, R., Wolfe, K.H., O'Byrne, C.P., Morrissey, J.P., 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. Antonie Van Leeuwenhoek 100, 507–19. doi:10.1007/s10482-011-9606-x

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. 24, 17–26. doi:10.1016/j.fbr.2010.01.001

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–8. doi:10.1093/bioinformatics/btu531

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47. doi:10.1186/s13068-015-0227-x

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–89. doi:10.1101/gr.1224503

Morrissey, J.P., Etschmann, M.M.W., Schrader, J., Billerbeck, G.M., 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. Yeast 32, 3–16. doi:10.1002/YEA.3054

Nonklang, S., Abdel-Banat, B.M.A., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M., Akada, R., 2008. High-Temperature Ethanol Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. Appl. Environ. Microbiol. 74, 7514–7521. doi:10.1128/AEM.01854-08

Nonklang, S., Ano, A., Abdel-Banat, B.M.A., Saito, Y., Hoshida, H., Akada, R., 2009. Construction of flocculent *Kluyveromyces marxianus* strains suitable for high-temperature ethanol fermentation. Biosci. Biotechnol. Biochem. 73, 1090–5.

Ozcan, S., Johnston, M., 1999. Function and regulation of yeast hexose transporters. Microbiol. Mol. Biol. Rev. 63, 554–69.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Proux-Wéra, E., Armisén, D., Byrne, K.P., Wolfe, K.H., 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics 13, 237. doi:10.1186/1471-2105-13-237

Ribeiro, O., Gombert, A.K., Teixeira, J.A., Domingues, L., 2007. Application of the Cre-loxP system for multiple gene disruption in the yeast *Kluyveromyces marxianus*. J. Biotechnol. 131, 20–26. doi:10.1016/j.jbiotec.2007.05.027

Riley, M.I., Sreekrishna, K., Bhairi, S., Dickson, R.C., 1987. Isolation and characterization of mutants of *Kluyveromyces lactis* defective in lactose transport. Mol. Gen. Genet. 208, 145–51.

Rocha, S.N., Abrahão-Neto, J., Gombert, A.K., 2011. Physiological diversity within the *Kluyveromyces marxianus* species. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 100, 619–630. doi:10.1007/s10482-011-9617-7

Rodicio, R., Heinisch, J.J., Heinisch, J.J., 2013. Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. doi:10.1002/yea.2954

Rubio-Texeira, M., 2006. Endless versatility in the biotechnological applications of *Kluyveromyces* LAC genes. Biotechnol. Adv. 24, 212–225. doi:10.1016/j.biotechadv.2005.10.001

Singh, M. V, Weil, P.A., 2002. A method for plasmid purification directly from yeast. Anal. Biochem. 307, 13–7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Tschopp, J.F., Emr, S.D., Field, C., Schekman, R., 1986. GAL2 Codes for a Membrane-Bound Subunit of the Galactose Permease in *Saccharomyces cerevisiae.* J. Bacteriol. 166, 313–318.

Varela, M.F., Brooker, R.J., Wilson, T.H., 1997. Lactose carrier mutants of *Escherichia coli* with changes in sugar recognition (lactose versus melibiose). J. Bacteriol. 179, 5570–3.

Wang, C.-J., Bajpai, R.K., 1997. A mathematical model of ethanol fermentation from cheese whey. Appl. Biochem. Biotechnol. 63–65, 495–509. doi:10.1007/BF02920449

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P., Boles, E., 1999. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. FEBS Lett. 464, 123–8.

Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., Sutthikhumpha, S., Pukahuta, C., Kitagawa, T., Nakamura, M., Cha-aim, K., Limtong, S., Hoshida, H., Akada, R., 2013. Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. Yeast 30, 485–500. doi:10.1002/yea.2985

Figure legends

Figure 1. Lactose consumption is a variable trait among *K. marxianus* strains. A set of 15 strains was grown in YPLac media. Growth (open bars) and lactose consumed (solid bars) were evaluated after 11 hours. Lactose consumption was calculated as the difference between the initial and final lactose concentration in the medium. Experiments were performed in triplicates with error bars showing SD.

Figure 2. Growth, consumption and transport of glucose and lactose by *K. marxianus* strains. CBS 397 (open circles) and CBS 6556 (solid circles) were grown in triplicate on MM medium supplemented with 2 % glucose (**A**), 2 % lactose (**B**) or 0.2 % lactose (**C**). A_{600} was monitored every hour and supernatants were analysed by HPLC. Sugar content in the medium is shown as white (CBS 397) and grey bars (CBS 6556). (**D**) β - galactosidase assays were performed in K. marxianus CBS 397 and CBS 6556 and in *K. lactis* CBS 2359. Cells were grown in YPD (black bars) and YPLac (white bars) to mid-exponential phase at which point the assay was carried out. (**E**) Sugar uptake assays were performed on cells pre-grown on 2 % glucose or lactose to an A_{600} of 2. Radiolabelled glucose (black bars) or lactose (white bars) was added and the kinetics of uptake measured. Histograms show the average ± S.D. of 3 independent experiments.

Figure 3. *K. marxianus* genome contains multiple copies of the lactose permease gene. The figure represents the CBS 6556 genome and was constructed with the information obtained from the CBS 6556 and DKMU 3 - 1042 genomes. Adjacent genes, shown in grey, are named according to their *S. cerevisiae* orthologues. *KmLAC12* and *KmLAC4* refer to the ancestral lactose permease and β-galactosidase genes, respectively.

The *LAC4-2* and *LAC12-4* pseudo-genes are shown as arrows containing asterisks that indicate mutations that interrupt the reading frame. Blue squares represent putative Gal4p binding sites. Distances between genes are not drawn to scale.

Figure 4. Growth on lactose by DMKU 3 -1042 x NCYC 1429 progenies. 190 random progenies were selected and grown in 96-well plates containing YPLac. Growth was measured after 24 hours using a multi-well spectrophotometer. Blue arrows indicate the parental strains DMKU 3 – 1042 and NCYC 1429.

Figure 5. Gene expression analysis of the *LAC12* gene copies. CBS 397 (black bars) and CBS 6556 (white bars) were grown in glucose, raffinose and lactose media. Then, RNA extraction and cDNA synthesis were carried out. Relative quantification of gene expression was performed by quantitative PCR. The values shown correspond to delta Ct calculations obtained when comparing against the expression of internal control *ACT1.* Histograms show the average ± S.D. of 3 independent experiments.

Figure 6. Functional analysis of the *LAC12* genes in *S. cerevisiae* EBY.VW4000 and *K. marxianus* DMKU3-1042. A. *S. cerevisiae* EBY.VW4000 strains expressing the different *LAC12* copies and the *LAC4* gene were grown in SC maltose media lacking leucine and uracil to an A_{600} of 2. Cells were collected by centrifugation and then washed twice with sterile water. Cultures were serially diluted (10-fold) and then inoculated into different carbon sources using a replicator. Plates were incubated for a period of 3-5 days at 30 °C. **B.** Overexpression of the *LAC12* genes in DMKU3–1042 *ura3-1* (RAK 3605). The strains expressing the *LAC12* genes were grown in YPLac medium. A_{600} was measured after 24 hours. The wild-type strains NCYC 1429 and DMKU3-1042 were used as controls for

efficient and inefficient lactose utilisation, respectively. Experiments were performed in triplicates with error bars showing SD.

Figure 7. Amino acid sequence alignment of Lac12p sequences from CBS 6556, CBS 397 and *K. lactis* CBS 2359. Differences between the sequences are represented by coloured residues; red-labelled residues represent amino acids conserved in efficient or non-efficient lactose-consuming strains whereas grey-labelled positions show changes between the CBS 397 and CBS 6556 sequences that are not necessarily conserved in other efficient / non-efficient strains. Blue rectangles represent transmembrane domains as predicted by InterProScan. The alignment was computed using MUSCLE.

Figure 8. Phylogeny of lactose transporters from *K.* marxianus. Maximum-likelihood phylogenetic tree generated by comparing amino acid sequences of lactose permeases and hexose transporters. KlLac12p: lactose permease from *K. lactis*, KmHgt1: putative galactose transporter from *K. marxianus*. ScHxt1 and ScGal2: low-affinity glucose transporter and galactose transporter from *S. cerevisiae.* The tree was created using MEGA. Numbers represent % bootstrap values based on 1000 samplings.

Supplementary material

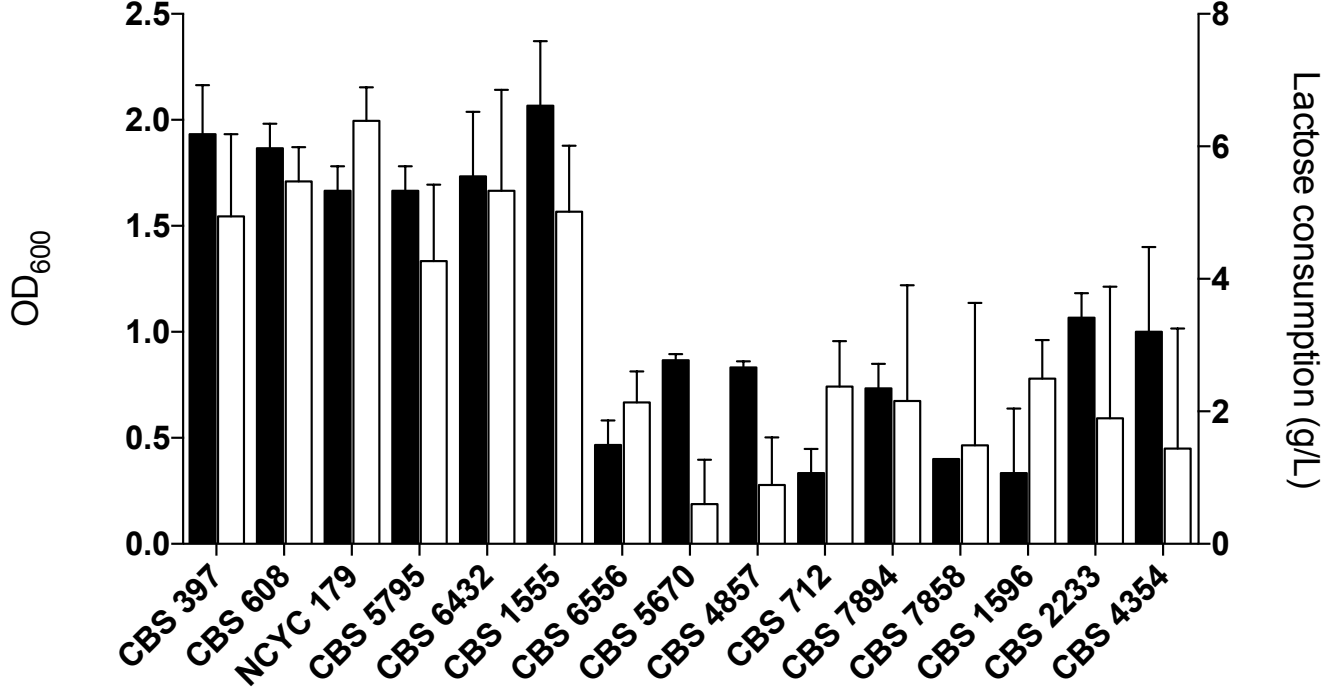
Figure S1. Plasmid YCp11256 use for overexpression of the *LAC12* genes. The *LAC12* genes were inserted between KmCDC19p and ScURA3 with substitution of yEmRFP. As result, the expression of the *LAC12* genes was driven by the *KmCDC19* promoter. The *KmCDC19p* sequence was obtained from *K. marxianus* DMKU3-1042. Sc*ADE2* and Sc*URA3* were obtained from conventional laboratory strains of *S. cerevisiae*. The plasmid replication elements, *KmARS7* and *KmCenD* were described by Hoshida et al., 2014.

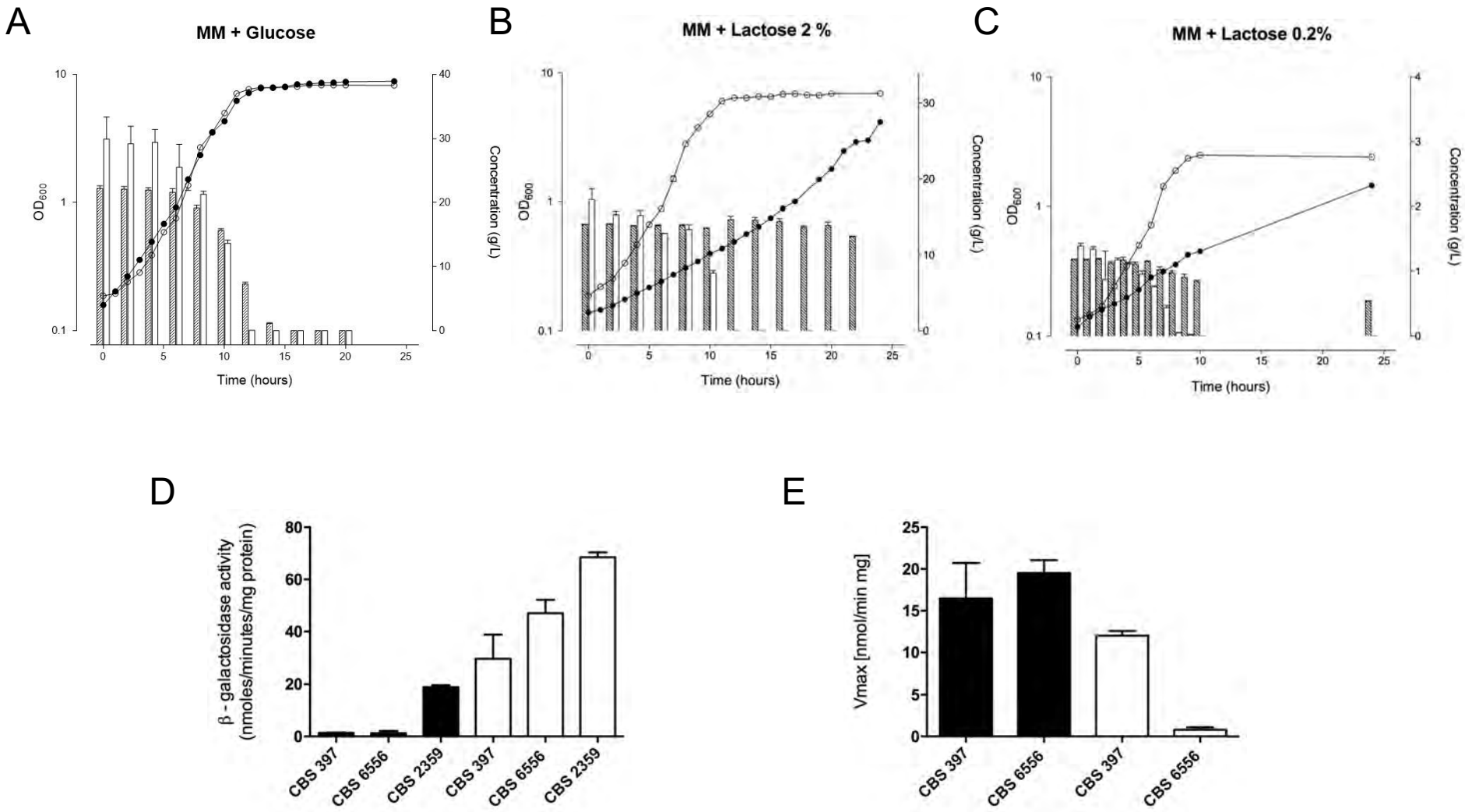
Figure S2. Presence of premature stop codons in the *LAC12-4* gene of some *K. marxianus* strains. The *LAC12-4* gene was sequenced and compared across all strains shown. Strains DMKU3-1042 and CBS 6556 contain mutations that independently cause premature stop in the sequence (red boxes). Numbers at the top of the figure correspond to coordinates of the *LAC12-4* gene in chromosome 8 of the DMKU 3 – 1042 genome.

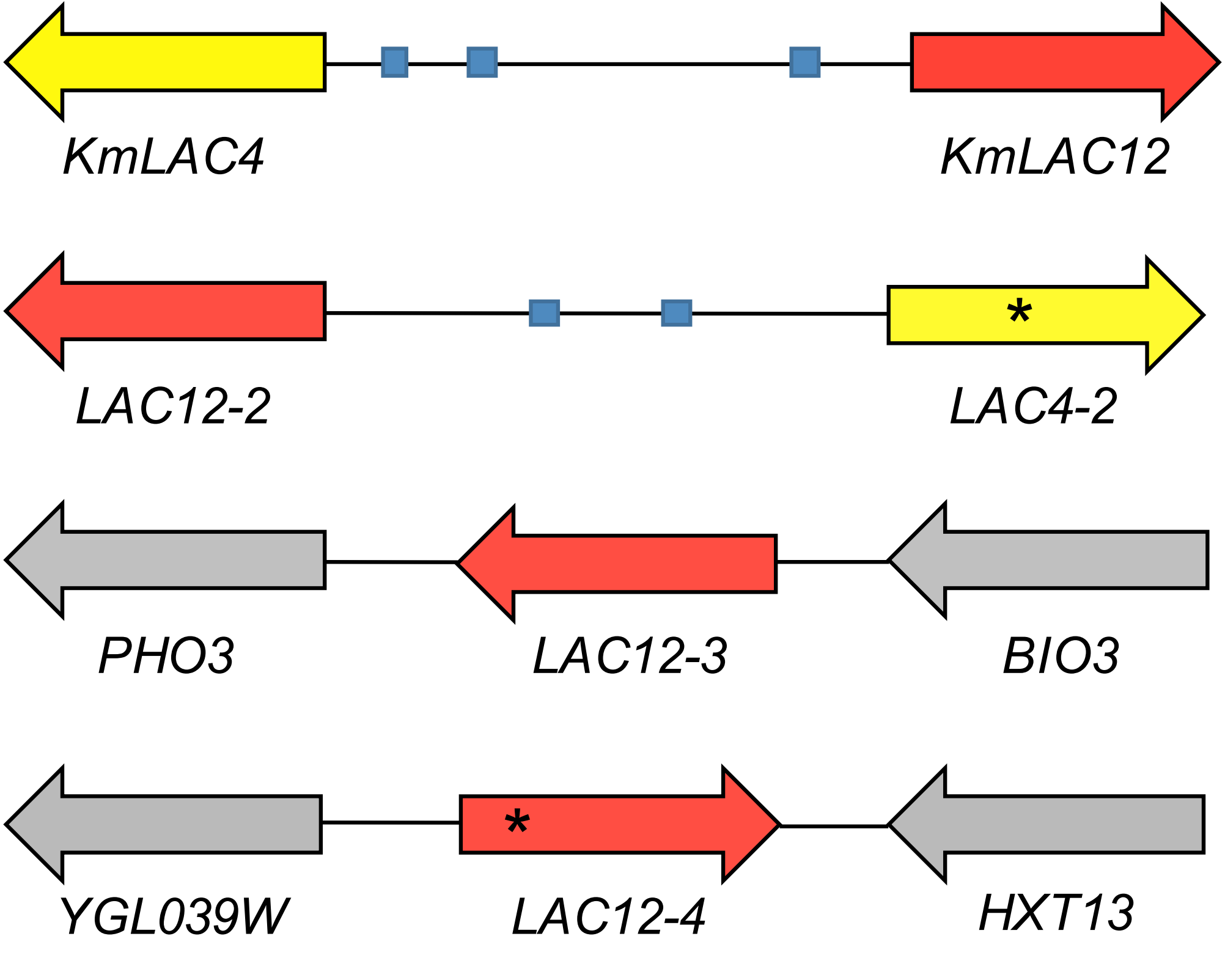
Figure S3. Growth of lactose medium by NCYC 1429 and DMKU3-1042 strains. Cells were grown in triplicates in a 96-well plate containing YPLac medium. Growth was monitored hourly. NCYC 1429 and DMKU3 – 1042 are shown in open and closed circles, respectively.

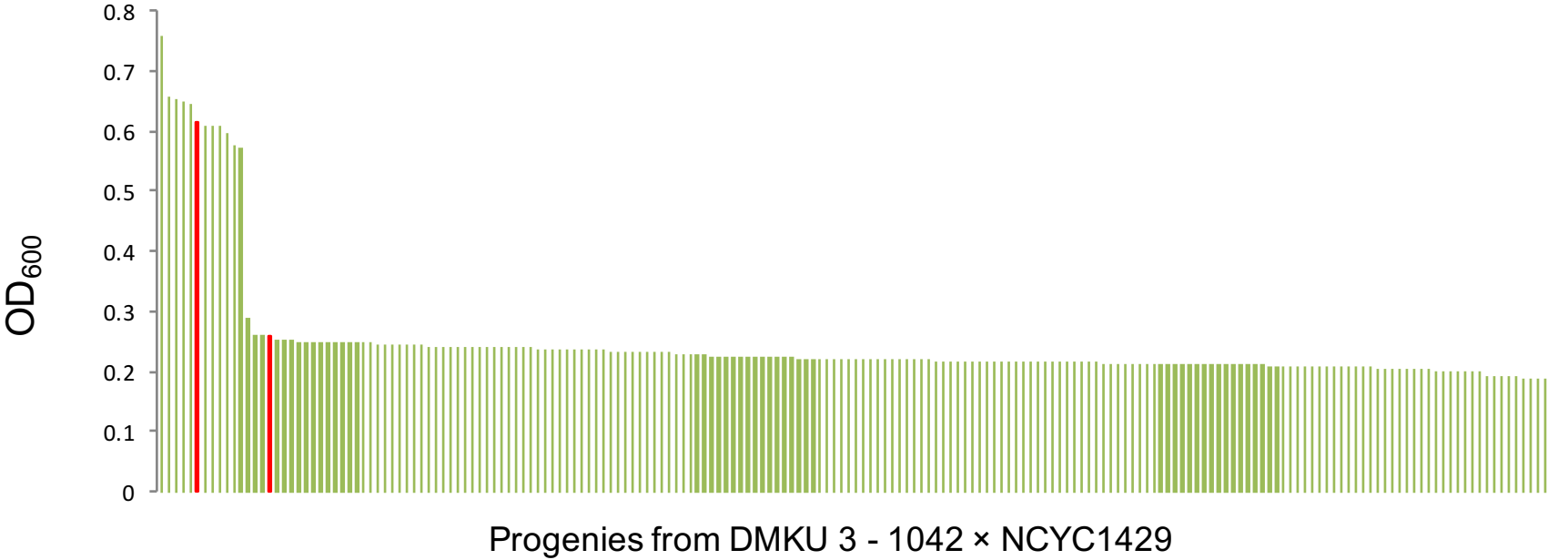
Figure S4. Subcellular location of the *Km*Lac12p - CBS 6556 transporter in *S. cerevisiae*. N-terminal GFP fusions were constructed for all the *LAC12* genes. Plasmids were transformed into *S. cerevisiae* BY4741, grown in presence of galactose, collected and visualized using a Zeiss LSM 5 laser scanning confocal microscope with a Plan-Apochromat 63x/1.40 Oil DIC M27 lens. **A.** Control strain expressing the GFP gene. **B.** Cells expressing the *KmLAC12::GFP* gene from CBS 6556. Similar results were obtained for *KmLAC12*, *LAC12-2* and *LAC12-4* – 397.

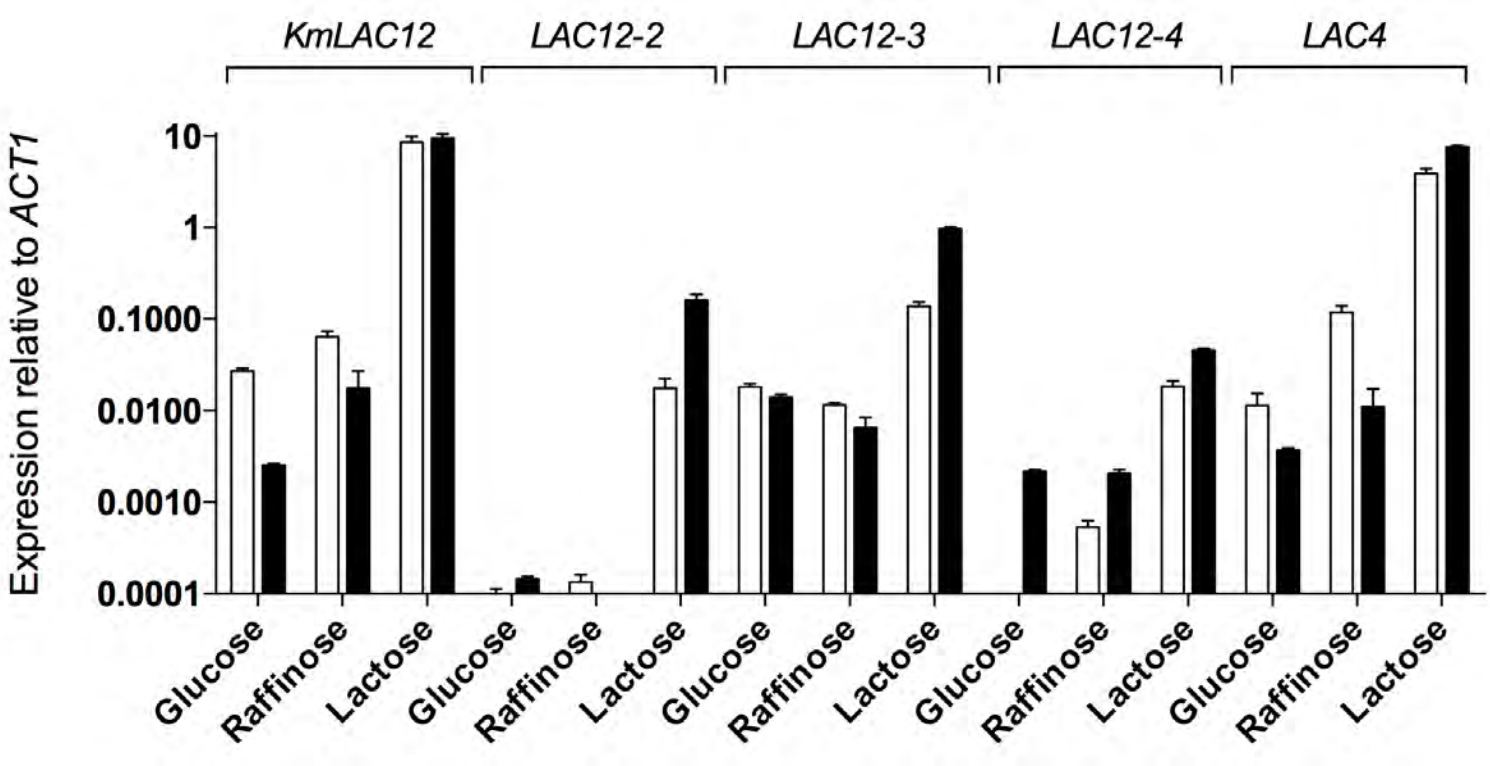
Figure S5. Multiple sequence alignment of *KmLAC12* sequences from *K. marxianus* strains. The *KmLAC12* gene was sequenced in 12 strains. Full-length amino acid sequences were obtained and used to generate a multiple sequence alignment in AliView. Strain names are shown at left. Residues that differ in good and poor utilisers are labelled in red.

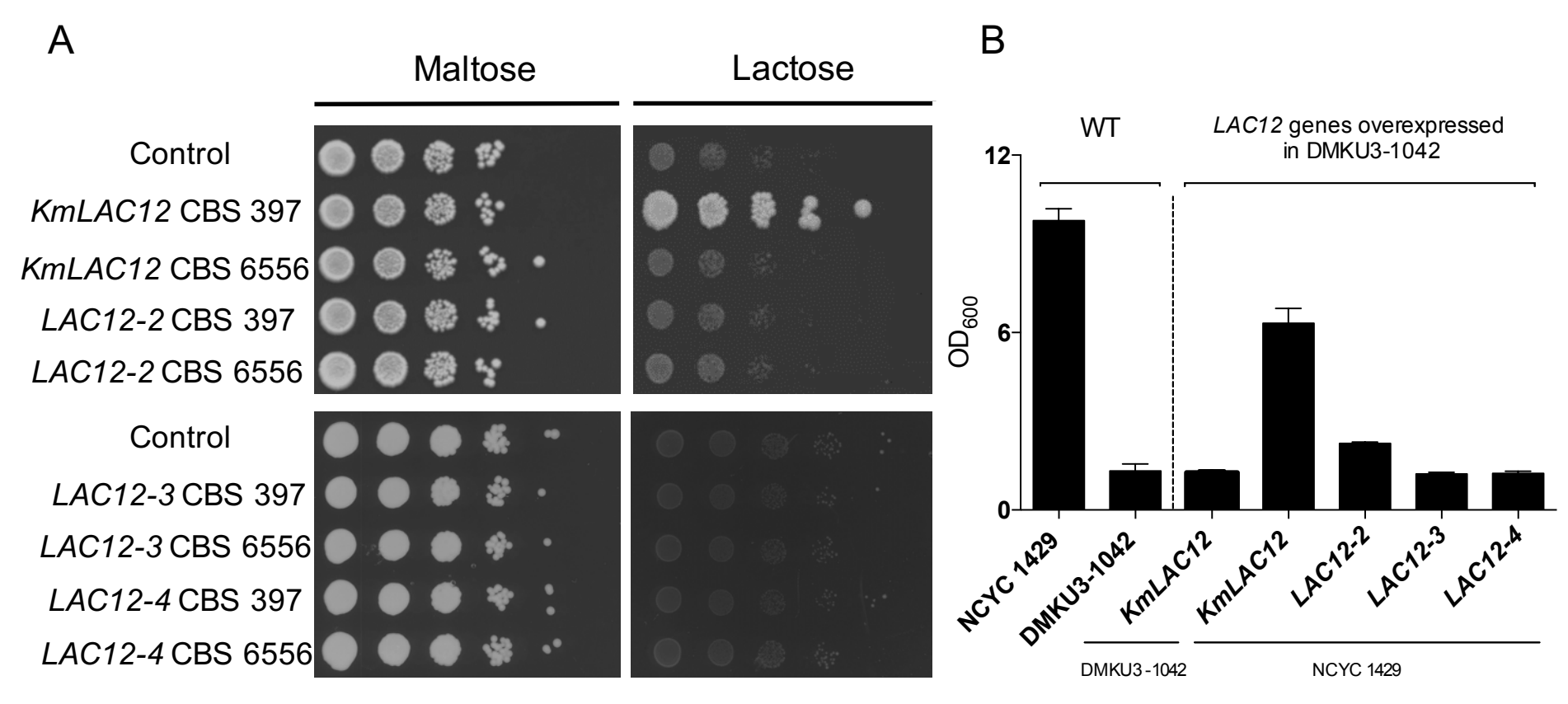

OD600
Lactose consumption (g/L)
2.5
2.0
1.5
1.0
0.5
0.0
8
6
4
2
0
CBS 397
CBS 608
NCYC 179
CBS 5795
CBS 6432
CBS 1555
CBS 6556
CBS 5670
CBS 4857
CBS 712
CBS 7894
CBS 7858
CBS 1596
CBS 2233
CBS 4354

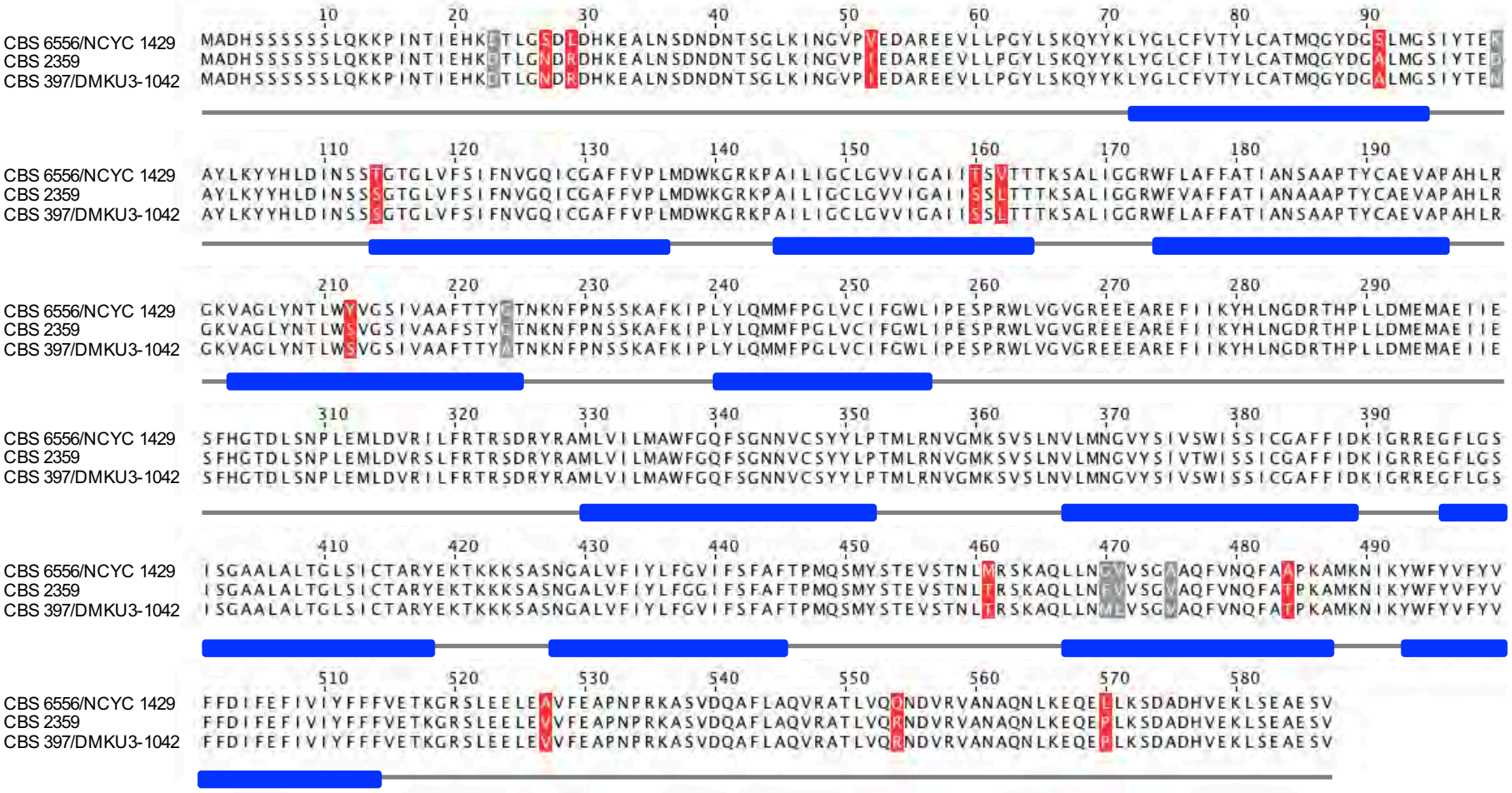

A
MM + Glucose
OD600
Concentration (g/L)
Time (hours)
B
MM + Lactose 2 %
OD600
Concentration (g/L)
Time (hours)
C
MM + Lactose 0.2%
OD600
Concentration (g/L)
Time (hours)
D
β - galactosidase activity (nmoles/minutes/mg protein)
CBS 397
CBS 6556
CBS 2359
CBS 397
CBS 6556
CBS 2359
E
Vmax [nmol/min mg]
CBS 397
CBS 6556
CBS 397
CBS 6556

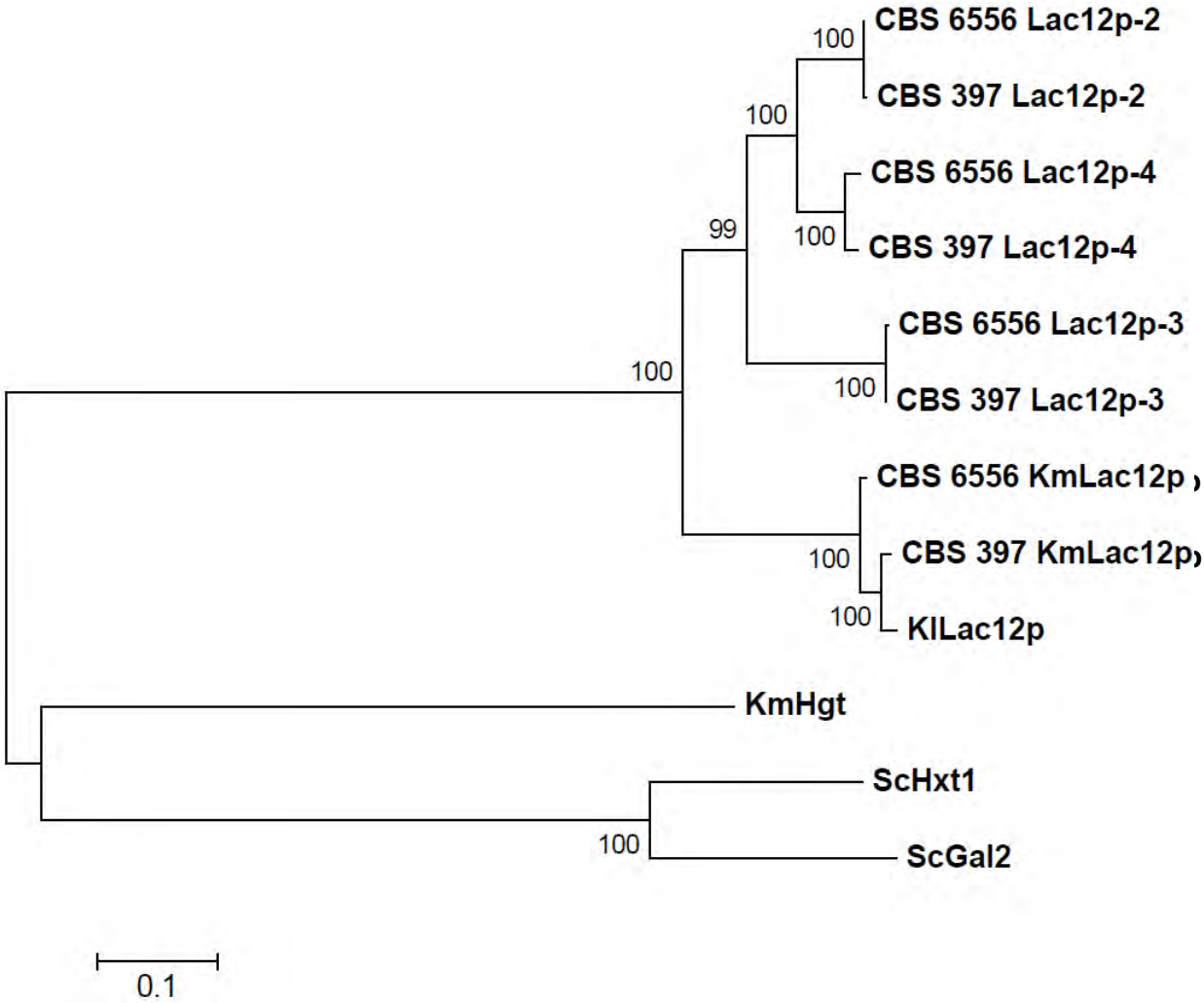

KmLAC4
KmLAC12
LAC12-2
*
LAC4-2
PHO3
LAC12-3
BIO3
YGL039W
*
LAC12-4
HXT13

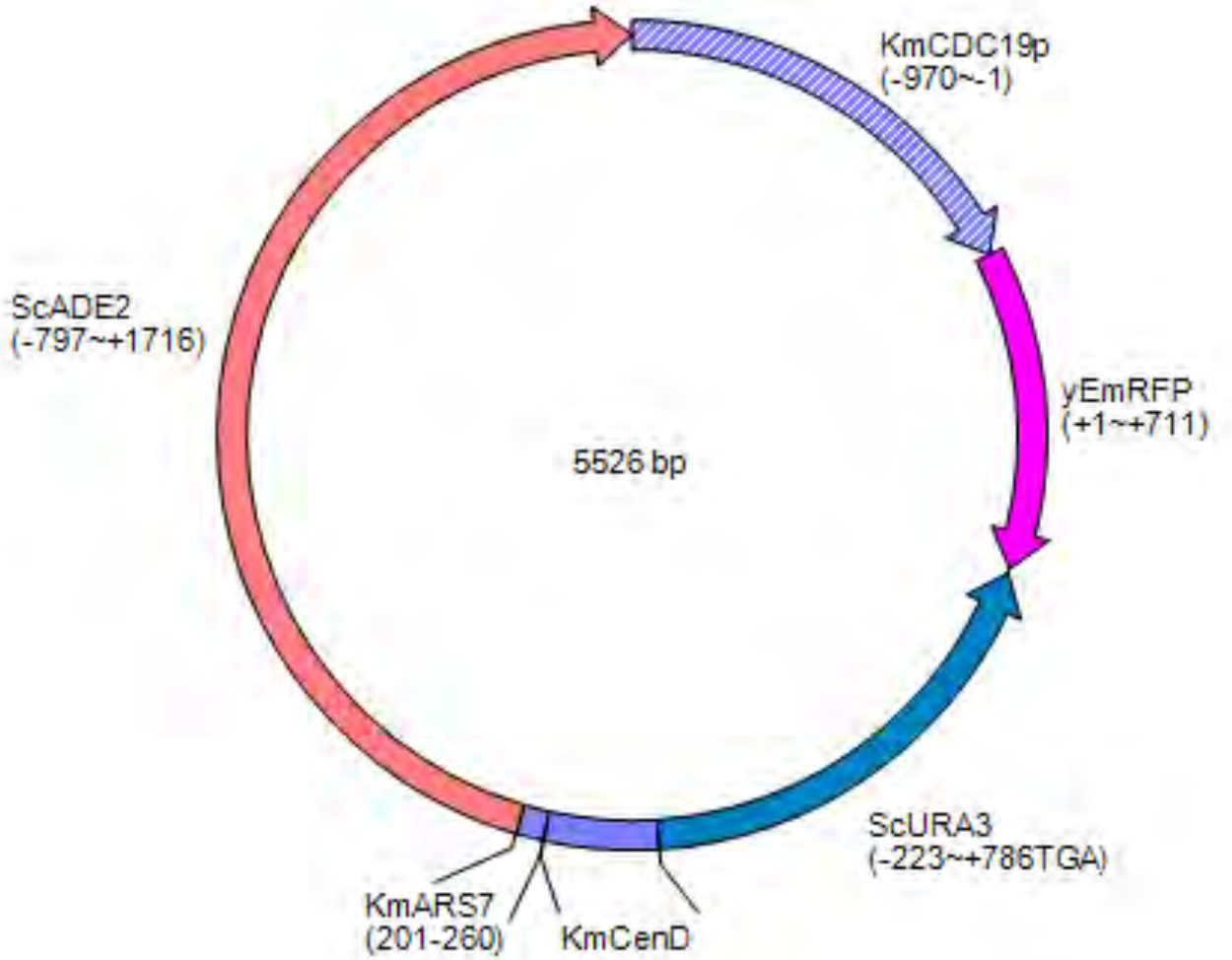

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
OD600
Progenies from DMKU 3 - 1042 × NCYC1429


KmLAC12
LAC12-2
LAC12-3
LAC12-4
LAC4
Expression relative to ACT1
10
1
0.1000
0.0100
0.0010
0.0001
Glucose
Raffinose
Lactose

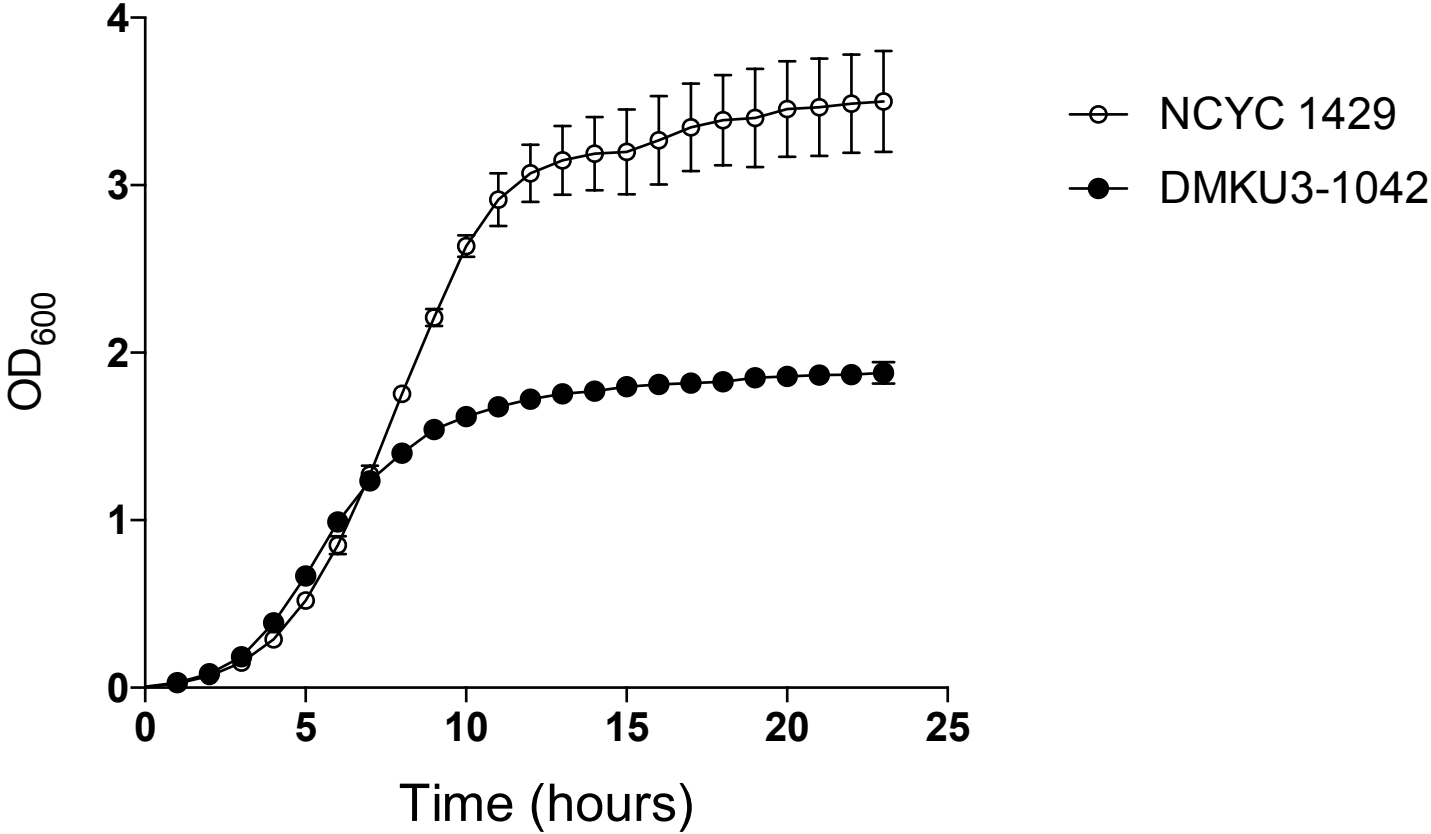

A
Maltose
Lactose
Control
KmLAC12 CBS 397
KmLAC12 CBS 6556
LAC12-2 CBS 397
LAC12-2 CBS 6556
Control
LAC12-3 CBS 397
LAC12-3 CBS 6556
LAC12-4 CBS 397
LAC12-4 CBS 6556
B
WT
LAC12 genes overexpressed in DMKU3-1042
12
6
0
OD600
NCYC 1429
DMKU3-1042
KmLAC12
KmLAC12
LAC12-2
LAC12-3
LAC12-4
DMKU3-1042
NCYC 1429

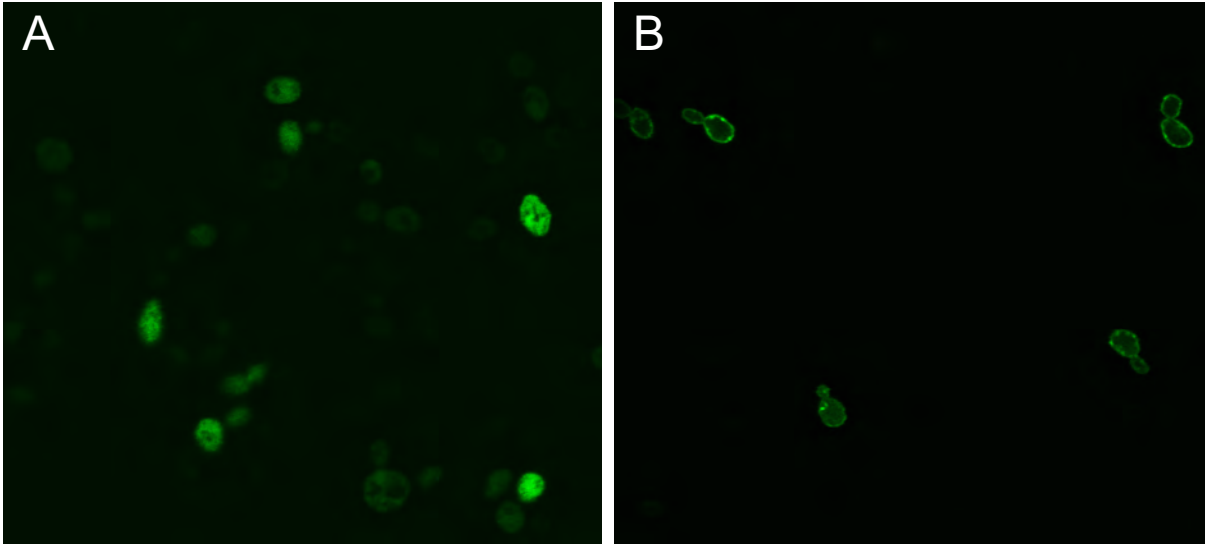

CBS 6556/NCYC 1429
CBS 2359
CBS 397/DMKU3-1042
MADHSSSSSSLQKKPINTIEHKDTLGSDLDHKEALNSDNDNTSGLKINGVPVEDAREEVLLPGYLSKQYYKLYGLCFVTYLCATMQGYDGSLMGSIYTEK
MADHSSSSSSLQKKPINTIEHKDTLGNDRDHKEALNSDNDNTSGLKINGVPIEDAREEVLLPGYLSKQYYKLYGLCFITYLCATMQGYDGALMGSIYTED
MADHSSSSSSLQKKPINTIEHKDTLGNDRDHKEALNSDNDNTSGLKINGVPIEDAREEVLLPGYLSKQYYKLYGLCFVTYLCATMQGYDGALMGSIYTEN
AYLKYYHLDINSSTGTGLVFSIFNVGQICGAFFVPLMDWKGRKPAILIGCLGVVIGAIITSVTTTKSALIGGRWFLAFFATIANSAAPTYCAEVAPAHLR
AYLKYYHLDINSSSGTGLVFSIFNVGQICGAFFVPLMDWKGRKPAILIGCLGVVIGAIISSLTTTKSALIGGRWFVAFFATIANAAAPTYCAEVAPAHLR
AYLKYYHLDINSSSGTGLVFSIFNVGQICGAFFVPLMDWKGRKPAILIGCLGVVIGAIISSLTTTKSALIGGRWFLAFFATIANSAAPTYCAEVAPAHLR
GKVAGLYNTLWYVGSIVAAFTTYGTNKNFPNSSKAFKIPLYLQMMFPGLVCIFGWLIPESPRWLVGVGREEEARERFIIKYHLNGDRTHPLLDMEMAEIIE
GKVAGLYNTLWSVGSIVAAFSTYGTNKNFPNSSKAFKIPLYLQMMFPGLVCIFGWLIPESPRWLVGVGREEEARERFIIKYHLNGDRTHPLLDMEMAEIIE
GKVAGLYNTLWSVGSIVAAFTTYATNKNFPNSSKAFKIPLYLQMMFPGLVCIFGWLIPESPRWLVGVGREEEARERFIIKYHLNGDRTHPLLDMEMAEIIE
SFHGTDLSNPLEMLDVRILFRTRSDRYRAMLVILMAWFGQFSGNNVCSYYLPTMLRNVGMKSVSLNVLMNGVYSIVSWISSICGAFFIDKIGRREGFLGS
SFHGTDLSNPLEMLDVRSLFRTRSDRYRAMLVILMAWFGQFSGNNVCSYYLPTMLRNVGMKSVSLNVLMNGVYSIVTWISSICGAFFIDKIGRREGFLGS
SFHGTDLSNPLEMLDVRILFRTRSDRYRAMLVILMAWFGQFSGNNVCSYYLPTMLRNVGMKSVSLNVLMNGVYSIVSWISSICGAFFIDKIGRREGFLGS
ISGAALALTGLSICTARYEKTKKKSASNGALVFIYLFGVIFSFAFTPMQSMYSTEVSTNLMRSKAQLLNGVVSGAAQFVNQFAAPKAMKNIKYWFYVFYV
ISGAALALTGLSICTARYEKTKKKSASNGALVFIYLFGGIFSFAFTPMQSMYSTEVSTNLTRSKAQLLNFVVSGVAQFVNQFATPKAMKNIKYWFYVFYV
ISGAALALTGLSICTARYEKTKKKSASNGALVFIYLFGVIFSFAFTPMQSMYSTEVSTNLTRSKAQLLNMLVSGVAQFVNQFATPKAMKNIKYWFYVFYV
FFDIFEFIVIYFFFVETKGRSLEELEAVFEAPNPRKASVDQAFLAQVRATLVQQNDVRVANAQNLKEQELLKSDADHVEKLSEAESV
FFDIFEFIVIYFFFVETKGRSLEELEVVFEAPNPRKASVDQAFLAQVRATLVQQNDVRVANAQNLKEQEPLKSDADHVEKLSEAESV
FFDIFEFIVIYFFFVETKGRSLEELEVVFEAPNPRKASVDQAFLAQVRATLVQQNDVRVANAQNLKEQEPLKSDADHVEKLSEAESV


CBS 6556 Lac12p-2
CBS 397 Lac12p-2
CBS 6556 Lac12p-4
CBS 397 Lac12p-4
CBS 6556 Lac12p-3
CBS 397 Lac12p-3
CBS 6556 KmLac12p
CBS 397 KmLac12p
KlLac12p
KmHgt
ScHxt1
ScGal2
100
100
100
99
100
100
100
100
100
100
0.1


KmCDC19p
(-970~-1)
yEmRFP
(+1~+711)
ScURA3
(-223~+786TGA)
KmCenD
KmARS7
(201-260)
ScADE2
(-797~+1716)
5526 bp


19,066
19,117
DMKU3-1042 TACTTAAAAGGCGTATTTGAACTATTATCATTTAGATGTAT-TTCATCAACA
CBS 6556 TACTTAAAAGGCGTATTTGAACTATTATCATTTAGATGTAT-TTCATCAACA
CBS 397 TACTCAAAAGGCGTATTTGAACTATTATCATCTAGATGTAAATTCATCAACA
CBS 608 TACTCAAAAGGCGTATTTGAACTATTATCATTTAGATGTAAATTCATCAACA
NCYC 111 TACTCAAAAGGCGTATTTGAACTATTATCATCTAGATGTAAATTCATCAACA
CBS 712 TACTCAAAAGGCGTATTTGAACTATTATCATCTAGATGTAAATTCATCAACA
CBS 7858 TACTCAAAAGGCGTATTTGAACTATTATCATCTAGATGTAAATTCATCAACA


4
3
2
1
0
OD600
0
5
10
15
20
25
Time (hours)
NCYC 1429
DMKU3-1042


A
B